The Language of Statistics

KEYWORDS:
Teaching;
Language.

Margaret Rangecroft
Sheffield Hallam University, England.
e-mail: f.m.rangecroft@shu.ac.uk

Summary
The discussion of problems associated with the use of language, specifically vocabulary and symbolism, is extended from the teaching and learning of mathematics to the particular area of statistics.

◆ INTRODUCTION ◆

Communication is at the heart of statistics. The job of the statistician very often involves acting as a consultant, sometimes to external clients but more usually internally to colleagues. This will include negotiating with the client about what is required and what is possible, as well as reporting back on findings. The statistician needs to be fluent in both the specialist language of the discipline and the everyday language understood by the client, and to be able to translate in both directions.

In much the same way we, as teachers, need not only to communicate in both 'languages' but also to teach our students to read, write and speak 'statistics'. Mathematics teachers are advised to devise activities to encourage students to write about mathematics and to have discussions (Simmons 1993; Thompson and Rubenstein 2000). In the best statistics teaching (i.e. that which is closest to real statistics) this is unnecessary because discussion and reporting are an integral part of the work. The insistence upon the use of a statistics problem in the Project for the new GCSE schemes (UK examinations normally taken at age 16) is clear evidence of this natural link between communication and statistics. In this sense it is much easier to solve some of the language problems in statistics than in mathematics.

It has long been recognized that language is an important issue in the teaching and learning of mathematics (Shuard and Rothery 1984) and the discussion continues to date, as a review of the many journals for mathematics teachers will show (e.g. Mahandru 2000). As statistics uses so much of mathematics terminology, it is clear that this research has much of interest for statistics teachers (most of whom will also be mathematics teachers and therefore familiar with the issues).

What has not been so widely aired is the fact that there are additional issues related specifically to the teaching and learning of statistics. This is a very large topic, covering as it does vocabulary, syntax, diagrams and symbols. This article will be confined to a discussion of some of the vocabulary and symbols in statistics, in other words, of Statistical English (SE).

◆ MATHEMATICAL ENGLISH ◆

Kane et al. (1967) coined the term Mathematical English (ME) to distinguish the specialist vocabulary used in mathematics from Ordinary English (OE). Shuard and Rothery (1984) use three categories of words in ME:

(i) words which have the same meaning in ME as OE
(ii) words which have a meaning only in ME
(iii) words which occur in both OE and ME, but have a different meaning in ME from their meaning in OE.

Many of the problems which students experience with language are not simply in learning the new terminology specified in the second of these, difficult as this may be. More difficult is coping with those words which occur in both ME and OE but with slightly (or even completely) different meanings. For example 'factor' in number and 'range' in statistics have very different meanings in OE. In other cases the difference is much more subtle, as in the use of the word 'event' in probability as opposed to OE. These subtle differences may in

fact cause more confusion amongst learners simply because teachers do not think the difference is sufficient to be worth pointing out. A more detailed categorization of the types of potential problems can be found in Thompson and Rubenstein (2000).

It is worth mentioning here the very complex differences in meaning which can be attached to the word 'mean'. In OE, 'mean' can be either an adjective or a verb. The *Chambers Dictionary* (Schwartz 1993) lists four quite separate definitions and even within these there are subtle differences, including:

Adjectives:
'... small-minded; stingy ...'
'... intermediate; average ...'
Verbs:
'... to intend; to purpose ...'
'... to moan; to complain ...'

We will ignore the last of these as it is so rare, but the other three meanings are in common usage. Our students will already have had to sort out the difference between the remaining verb and the adjectives but now we introduce the noun. The noun has some connection with the second definition above, but we are trying to teach that it is NOT the same as the median as implied by the use of the word 'intermediate'. To make matters worse the student will, most likely, already have used the word 'average' to signify what we are now calling the mean. We add to the potential confusion by trying to get across the idea that there are several 'averages', one of which is the mean. Is it surprising that some students become confused?

◆ STATISTICAL ENGLISH ◆

The situation in statistics is further complicated by the fact that there are three more categories to add to Shuard and Rothery's list:

(i) words which have a meaning only in SE
(ii) words which occur in both OE and SE, but have a different meaning in SE from their meaning in OE
(iii) words which occur in both ME and SE, but have a different meaning in SE from their meaning in ME.

We all know of the particular jargon of statistics, words such as median, hypothesis and parameter. The problems of understanding their meanings are analogous to the problems of mathematical jargon.

The second case in our list above highlights the need to realize the fact that words such as 'confidence' and 'error' will already be familiar to our students in OE, but it may take some time for students to adapt to their alternative meaning in SE. However it is the last case in our list which is perhaps the most easily overlooked. Take, for example, the verb 'estimate'. In ME (as well as often in OE) this can be translated as 'to make an educated guess' or 'to approximate', whereas in SE it is used to describe the very precise calculation of the 'prediction' of the value of a parameter. A recent specimen GCSE examination paper (Edexcel 2000) used both of these meanings in different questions as follows:

7. There are four blood groups A, B, AB and O.
The probability that a person picked at random will be in blood group A, B or AB is shown in the table.

Blood Group	*A*	*B*	*AB*	*O*
Probability	0.42	0.09	0.03	

(a) Work out the probability that a person chosen at random will be in blood group O.

Mathstown High School has 1000 students.

(b) Work out an estimate for the number of students who will be in blood group A.

15. $$P = \frac{mv}{t}$$

$m = 324$, $v = 76.8$ and $t = 0.413$

(a) Work out an estimate for the value of P.

(b) Make v the subject of the formula $P = \frac{mv}{t}$

(Reproduced by kind permission of Edexcel)

As this paper did not allow the use of calculators, the students were expected to approximate in question 15(a). But the 'exact' value of 420 was required in answer to 7(b), although both questions gave the instruction to 'estimate'. Although the usage is quite correct in both questions, some of the students who completed this paper recently as a mock examination gave the answer 400 for 7(b). They had employed the same approach as in question 15, where it was sensible to approximate. This is hardly surprising given that exactly the same wording had been used in both questions. Interestingly, those pupils who made this mistake were at the bottom end of the ability range within an intermediate group.

This is not an isolated instance. Take, for example, the word 'significant', which has one meaning in OE, a different meaning as part of 'significant figures' in ME and yet another meaning in SE as

part of 'significant result'. You can probably think of many more.

◆ OVERCOMING THE PROBLEMS ◆

Knowing that a problem exists is the first step to 'solving' it. If as teachers we can become more attuned to the possibilities of misunderstandings arising from language difficulties, we can perhaps recognize them and make the necessary explanations. Maybe we should be taking one step back and trying to pre-empt difficulties by careful use of language in our own teaching. This sounds much easier than it is in reality. As relative experts in this area we take in our stride the subtle differences and are able to switch from OE to ME to SE without consciously thinking about them. So it is not surprising that we do not always tune in to the difficulties which our students have. It takes a lot of practice and sensitivity to students' needs if one is to become anywhere close to clear in all one's explanations. It is also important to be able to diagnose problems arising from language difficulties correctly.

To make our meaning clear to students, two conflicting but guiding principles can be employed:

Explain everything in OE and avoid ME and SE wherever possible

OR

Always use the correct vocabulary for the situation so as to make explanations crystal clear.

So, should we aim for accessibility or precision? Do we talk about the average because that is a term with which most children are familiar, or do we use the word mean which is strictly correct but has an alternative meaning (sorry!) in OE? There cannot be one answer to this dilemma as both approaches have their merits and drawbacks. It must be a professional judgement by the teacher based on experience of dealing with a particular age and ability group. Generally there has been a tendency in the UK to use OE in Primary Schools (times, average etc.) and ME and SE in Secondary Schools and Higher Education (multiplied by, mean etc). There is also some evidence that teachers at all levels use OE where possible in speaking to students, but the textbooks to which they refer use ME and SE. Are these strategies helpful or do they cause even more confusion?

◆ SYMBOLISM ◆

The use of symbols and specialist notation is common in many subject areas, particularly, but not exclusively, in the sciences. It is generally in mathematics and statistics that most students meet them first. Again the problems that they cause in mathematics are well documented and can be assumed to apply equally to statistical symbols and notation. But what about the symbols and notation which are found in both mathematics and statistics but have different meanings? For example, 'f' in mathematics usually means 'function', as it does sometimes in statistics, but where it commonly means 'frequency'. Can our students cope with such anomalies? Probably our more able students take all these things in their stride, but what about those who find them a stumbling block?

◆ CONCLUSION ◆

Statistics is greatly enriched by its close association with language, but this brings its own special problems. We have looked briefly at just some of the particular difficulties with vocabulary and symbolism because these are two areas where there are problems specific to statistics, but there are also the problems of syntax and readability that are shared with mathematics. There are even problems in statistics which do not fit into any of the above categories but where the approach is not what is expected in mathematics. A nice example of this can be found in volume 21 issue 3 of *Teaching Statistics* (Goodall 1999), where the Editor introduces an A-level (UK age 18) examination question in which it is necessary to disregard the convention of measuring angles anticlockwise from East as generally used by mathematicians.

My own interest in this area arose from working with a group of students for whom English was a second language. This highlighted for me the difficulties posed by language in teaching statistics, but I am convinced that the problems exist across the whole spectrum of backgrounds and abilities although they may not always be so obvious. I hope that my heightened awareness has helped me to be more sympathetic to students' difficulties. It is a great pity when potentially able students are held back by 'simple' language difficulties that can be overcome so easily if we recognize them.